1 Introduction

A fundamental prediction of biological evolution is that a species will most commonly share many characteristics with lineages from which it has recently diverged, and fewer characteristics with lineages from which it diverged further in the past. This principle, which results from descent with modification, is one of the most basic in biology (Darwin 1859). A reconstruction of the relationships between species that is based on similarities and differences in characteristics – a phylogenetic hypothesis – can be used as a powerful tool to understand fundamental questions about the history of life (Nunn 2011; Baum and Smith 2013; Harmon 2019).

Phylogenetic comparative methods (PCMs) can be broadly defined as statistical approaches that incorporate information about the shared evolutionary history of taxa (i.e. their nonindependence) to identify macroevolutionary patterns or test hypotheses about how those patterns relate to macroevolutionary drivers, such as climate or biotic interactions. Until recently, there have been largely separate analytical frameworks for phylogenetic inference (inferring relationships between taxonomic units using morphological or molecular character data) and PCMs (testing hypotheses about evolution while treating the relationships as known). In this Element we only discuss the latter, but other recent literature focuses on a unified methodological framework that integrates the two (Warnock and Wright 2020; Wright et al. 2020). Many other types of information, such as biogeographic data (Matzke and Wright 2016; Landis 2017), can be used in this modeling framework, and as new models are developed the scope of questions that can be addressed will increase.

There are many reasons (perhaps particularly for paleobiologists) why it might not be feasible, or of interest, to use this kind of unified framework, where you must infer phylogenetic relationships as a means to answer other macroevolutionary questions. Perhaps you have previously estimated a phylogeny but now want to use it to answer new questions, perhaps you are interested in combining several smaller phylogenies to generate a supertree, or perhaps your specific question is not yet answerable in a Bayesian process-based framework (e.g. Warnock and Wright 2020; Wright et al. 2020). The good news is that trees constructed in many different ways can be used in PCMs to make reliable inferences about trait diversification, provided that the tree is appropriately scaled to time using the stratigraphic record as an extra source of information (Bapst 2014a; Soul and Friedman 2015; Barido-Sottani, Tiel et al. 2020).

PCMs have been rapidly proliferating in the past five to ten years, and the kinds of questions they can be used to rigorously answer are now very diverse.

Here, we review and demonstrate some of the analytical approaches that can be applied when you already have a phylogeny in hand. Many of these PCMs are used to model trait change through time and the relationship between that trait change and other variables. Most do not model the underlying microevolutionary processes occurring in populations or the external drivers that generate the trait change. Instead, PCMs more commonly use stochastic models to investigate the long-term outcome of evolutionary change. Therefore, as is outlined in more detail later on, different underlying processes can generate similar patterns that can be equally well explained by the same model. Careful interpretation of the results of any analysis is imperative.

We begin by outlining some fundamental approaches that are conceptually important, and then move on to more complex macroevolutionary models. Multiple books have been written on this vast topic (see Nunn 2011, Garamszegi 2014, and Harmon 2019, each of which are not focused on the fossil record but paleontologists might find them useful nonetheless), but we hope to provide you with a clear, digestible explanation of the theory behind a variety of PCMs, along with enough information on how to apply them to get started in answering your own questions. A reader already well versed in these approaches will find a review of recently published methods, and suggestions for their implementation in a paleobiological context, in the last third of the Element. A variety of software has been used to implement PCMs, but the majority of those likely to be of interest to paleontologists are available in R (e.g. Bapst 2012; Lloyd 2016; Barido-Sottani et al. 2019). We therefore provide all of our reproducible examples in R. Package names and inline code examples are in `Courier New` font.

All data used in the examples in this Element are available on GitHub (https://github.com/daveyfwright/PCMsForPaleontologists), along with a script to load and format the data in the R programming environment so that they can be analyzed immediately, as well as the full R script and annotated script and outputs from the example analyses. Although the examples in this Element are intended to provide a guide for implementing comparative analyses in R, we encourage readers to also follow along directly using the scripts we provide on GitHub. The example code in the main Element assumes a basic working knowledge of R, including loading data, inspecting objects, reading help pages, and object assignment. The materials online do not make this assumption and provide a detailed step-by-step guide.

2 Getting Started: Data and Phylogeny

Functions in R that can be used to manipulate phylogeny or apply phylogenetic comparative approaches make use of trees that are in "`phylo`" format,

originally implemented in the ape package (Paradis et al. 2004). In this section, we outline features of this format, as well as some important things to remember when preparing paleontological data for an analysis in R (see also Bapst 2014b). The dataset we use here is for fossil crinoids (Eucladida, Echinodermata), a morphologically diverse clade of marine invertebrates with a well-sampled fossil record. Our use of this dataset is primarily intended to demonstrate the different tree-based analytical tools that can be applied, rather than to glean specific inferences about crinoid macroevolution, and we caution readers that our results should be viewed in this light.

Key packages in R that contain implementations of PCMs that are commonly applied to fossil data are `ape` (standard format and processing for phylogenies in R – Paradis et al. 2004), `nlme` (fitting Gaussian models, e.g. least-squares regression – Pinheiro et al. 2019), `geiger` (a versatile package that performs and plots many PCMs – Harmon et al. 2008; Pennell et al. 2014), `phytools` (additional plotting and simulation functions – Revell 2012), and `OUwie` (heterogeneous macroevolutionary model fitting, e.g. Brownian motion or Early Burst – Beaulieu and O'Meara 2020). There are many others, so it is valuable to spend time exploring the different tools that might be best suited to answering a particular question. Once you have gained familiarity with the above key packages, a very extensive list of all the packages that can be used for phylogenetic approaches in R can be found on this website: https://cran.r-project.org/web/views/Phylogenetics.html.

2.1 Phylogeny

For most of the example analyses we use a single phylogeny of eucladid crinoids that has branch lengths that represent time in millions of years (Figure 1, called `tree` in our scripts). It is the maximum clade credibility tree (MCCT), inferred using a model of morphological character evolution combined with a process-based model of diversification that allows inference of ancestor–descendant relationships (this method is distinct from ancestral *state* reconstruction – see subsection 4.2; Wright 2017a). The MCCT is the tree in the posterior distribution with the largest product of clade frequencies (i.e. probabilities), which represents a point estimate of phylogeny in a Bayesian context. Our emphasis on a single tree is for illustrative purposes only; in reality, analyses should *always* be applied to a set of possible phylogenies to assess how robust results are to variation in tree topology and branch lengths (see Section 6). The set of possible phylogenies (often referred to as a tree set) could be a set of the most parsimonious trees, a random sample from a Bayesian posterior distribution, repeats of stochastic

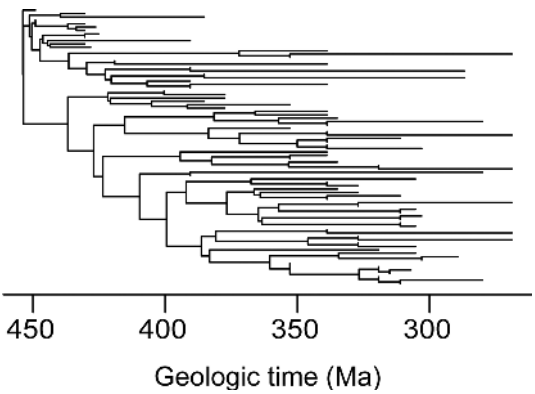


Figure 1 Phylogenetic tree comprising eighty-two species belonging to the Eucladida (Crinoidea, Echinodermata). Tree topology and branch lengths correspond to the maximum clade credibility tree (MCCT) based on a Bayesian tip dating analysis of morphological character data presented in Wright (2017a). Branch lengths are scaled to absolute time in units of millions of years. The MCCT is called `tree` in our scripts and represents a point estimate of phylogeny. Where possible, we recommend testing macroevolutionary inferences across a distribution of trees rather than a point estimate, and our emphasis on the MCCT is for illustrative purposes only (see subsection 2.1 and Section 6) .

timescaling of a composite topology (Bapst 2014a), etc. The number of trees to include in the set depends on how variable results are across different trees; if results are highly variable, a larger set will be needed to characterize the possible outcomes of the analysis and how common they are. For example, 100, 500, or 1,000 are common set sizes in the published literature. The phylogenies used for analysis should have branch lengths that are scaled to time (best practice would be to use a birth-death-sampling approach [Stadler 2010; Gavryushkina et al. 2014; Heath et al. 2014; Wright 2017; Stadler et al. 2018], but if that is not possible, see Hedman 2010, Bapst 2014b, Halliday and Goswami 2016, or Lloyd 2016 for options and considerations when choosing an a posteriori timescaling approach).

A tree in `phylo` format has branches that are called “edges” and branching points that are called “nodes”. For timescaled phylogenies, the branch (or edge) is a graphical representation of the amount of time since the lineage leading to one taxon diverged from its sister lineage, so the length of terminal branches does not represent the amount of time the actual species or genus at the tip was extant, only the time since divergence. Each node has a number assigned to it, beginning with the tips. This format is for both ultrametric (branches all end at

the same time; usually trees of all living taxa) and nonultrametric trees (branches end at different times; usually trees that include extinct taxa).

A `phylo` tree in R has four components: (1) a matrix that identifies how the branches connect, by listing their start and end node numbers, (2) a vector of tip labels (user defined, usually taxon names), (3) a vector of branch lengths, and (4) an integer number that is the number of internal nodes. For trees that include only extinct taxa, the root age of the tree can be set using `tree$root.time <- X`, where X is the numerical age (usually reported in units Ma for paleontological datasets). A root age is required by some analyses, and facilitates good visualization when plotting a tree. If there is no root age assigned, most functions will assume the youngest tip ends at the present day; without it, the tree is in relative time, rather than absolute.

An important first step prior to analysis is to plot the tree (Revell et al. 2018). In our eucladid example some taxa were inferred to be ancestral to others in the tree (Figure 1; see subsection 4.2 for clarification of what this means). In R, a `phylo` object displays these sampled ancestors as sister to their descendants, but with zero-length branches (i.e. no inferred change between the node and tip for the ancestral taxon). For many PCMs, zero-length branches are mathematically intractable (the reasons relate to division-by-zero issues). Adding a very short length to each zero-length branch resolves this problem. If you are concerned about the possibility of this introducing a bias in your own analysis, you could also drop these tips from the tree and compare the results using each tree. Note that in the Eucladida dataset there are many inferred sampled ancestors, so dropping these tips may represent high information loss. Whether or not this is the case for different datasets will depend on the group under investigation. Node labels are required by some PCM packages (e.g. `OUwie`), so it is best practice to assign them using a vector the same length as the number of nodes (as long as you check that the R function you are using doesn't use them for anything you aren't expecting; for example, `OUwie` uses them to define ancestral macroevolutionary regimes – see subsection 9.1). This can be done using `tree$node.label <- rep(1, Nnode(tree))`, which gives all nodes the label "1".

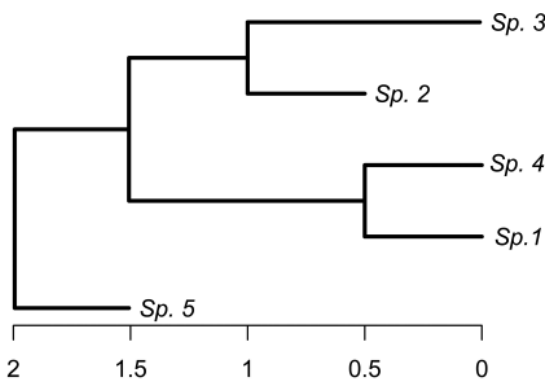
A very useful basic function to inspect and manipulate phylogenies in R is `vcv`. When applied to a timescaled tree this function outputs the phylogenetic variance-covariance (VCV) matrix of that tree. The VCV matrix is an intuitive numeric representation of the tree, and is used in the inner workings of many R functions implementing PCMs. Models of continuous trait change (like Brownian motion; see Section 4 onwards) differ from one another with respect to how the variance and covariance of the trait are expected to change through time, and the VCV matrix is the basis for the statistical expectation under

different models. Elements on the diagonal of the matrix give the variance. When a tree has branch lengths scaled to time (as paleontological trees usually do), these values give the root-to-tip distance for each taxon on the tree (i.e. duration). The functions `max(diag(vcv(tree)))` will output the maximum root-to-tip distance (i.e. the duration of the whole tree). The off-diagonal elements give the covariance, which is the amount of shared variance between pairs of taxa (i.e. the length of their shared evolutionary history). Figure 2 shows a small example tree and its associated VCV matrix.

2.2 Trait Data

In our example analyses we use a variety of PCMs to explore patterns in two continuous traits measured for each species in the Eucladida tree. These are `Shape` (calyx shape; the natural log of the length/width ratio of the calyx) and

a



b

	Sp. 1	*Sp. 2*	*Sp. 3*	*Sp. 4*	*Sp. 5*
Sp. 1	2.0	0.5	0.5	1.5	0.0
Sp. 2	0.5	1.5	1.0	0.5	0.0
Sp. 3	0.5	1.0	2.0	0.5	0.0
Sp. 4	1.5	0.5	0.5	2.0	0.0
Sp. 5	0.0	0.0	0.0	0.0	0.5

Figure 2 (a) Hypothetical phylogeny with branch lengths scaled to absolute time. (b) Variance-covariance matrix that describes the tree. The values along the diagonal of the matrix represent the amount of time from the root to each tip in the tree, which corresponds to their expected variance in a Brownian motion model. In contrast, the off-diagonal values represent the shared amount of evolutionary history for pairs of tips, which corresponds to their evolutionary covariance. Note that the diagonal values in the matrix represent terminal branches whereas the off-diagonal values represent internal branches.

`Density` (filtration fan density; the natural log of the approximate number of proximal feeding appendages an individual of the species has). We also briefly demonstrate analysis of a discrete trait `Complexity` (calyx complexity; the number of plates interrupting the posterior interray). We store all the traits in an object called `alltraits` for ease of use in R. Some analyses implemented in R require that the tree and data have exactly the same taxa; for these we use a tree (called `prunedTree` in code examples) in which taxa not present in the trait data have been removed from the tree. We use an estimated standard error of 0.03 for `Shape` and 0.12 for `Density`, based on an average across species for which there is more than one specimen.

3 Phylogenetic Nonindependence

Felsenstein (1985) was the first to outline the problem of the nonindependence of species trait data, and an algorithm to account for that problem, which he called phylogenetic independent contrasts (PIC). The most common evolutionary question that this problem (and proposed solution) is applied to involves the relationship between two traits. The extreme case of the problem of nonindependence of species data is shown in the original paper (Felsenstein 1985, figure 1), and a related example is provided in Figure 3 based on Nunn and Barton (2001). An early divergence within a clade leads to two groups of taxa that have quite different values for morphological traits between those two groups, and more similar trait values within each group, for both of the traits under investigation. When a linear regression model or correlation test is used, this will result in a strong relationship being inferred, but effectively a line is being fit to two points – the two groups within the clade, and no such strong relationship in fact exists. The early split means that species in the two parts of the tree have been evolving separately from one another for a long time, and so have had a long time to accumulate differences between them; species in the same part of the tree are more similar to each other because of their recent common ancestor and long shared evolutionary history. Regression analysis assumes that individual data points are statistically independent from one another; this assumption is violated by species data because of the shared evolutionary history.

PIC analysis is an intuitive approach to this problem. It takes the average trait value of sister clades and weights this by the amount of time since their most recent common ancestor (i.e. the amount of time they have been evolving separately). Most researchers now use phylogenetic generalized least squares (PGLS) instead of PIC because it is a more flexible likelihood-based model framework (PIC can be shown to be a special case of PGLS; see Garland and Ives 2000; Blomberg et al. 2012). After a nonphylogenetic ordinary least-

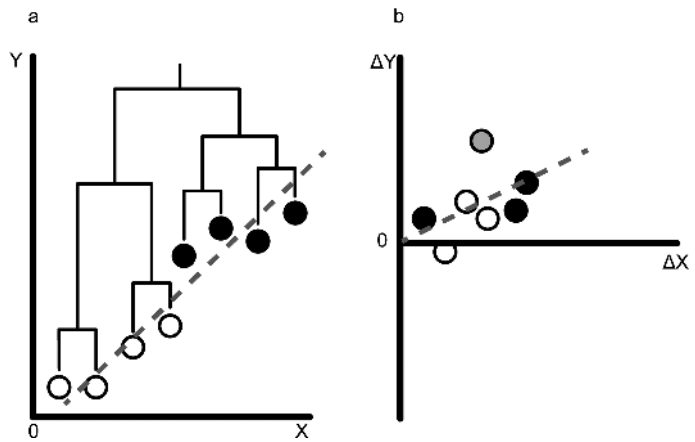


Figure 3 How species nonindependence may influence trait correlations and how phylogenetic independent contrasts address the issue. (a) Raw values for two traits (X and Y) plotted alongside the underlying phylogenetic relationships among species. Two subclades are identified and labeled by open and filled circles, corresponding to species with smaller (open) vs. larger (filled) trait values. The dashed line represents a best-fit linear model using ordinary least-squares (OLS) regression. (b) Phylogenetic independent contrasts estimated as the standardized difference between species trait values for each internal node in the tree. Note that the y-intercept goes through the origin, and the difference in slope. The contrast with the largest absolute value (gray circle) represents the evolutionary shift in trait values between the two subclades. Adapted from Nunn and Barton (2001).

squares (OLS) regression is performed on species data, there will be a phylogenetic signal that dictates how far each datapoint is from the regression line (its residual). We say therefore that there is phylogenetic structure in the residuals of the regression. PGLS incorporates information from the phylogeny into the regression model to adjust the regression line so that the residuals are normally distributed, rather than structured, rendering the analysis statistically valid.

Fitting both an OLS and PGLS regression line can be done with the same R function, `gls`. Generalized least-squares regression is a standard approach to regression that can be used when the datapoints are not independent from one another. When this is the case, the residuals of a standard regression will have a structure that is caused by the relationship between the datapoints. Generalized least-squares regression is a very flexible framework that allows you to supply a correlation structure for the residuals of the regression that can

be derived from any potential source of nonindependence. For example, here we are using phylogeny, so we supply the phylogenetic VCV matrix to define the expected structure in the residuals, but, in another example, because some traits are known to vary systematically across space (e.g. Wagner and Marcot 2010), a matrix of expected spatial covariance could also be used.

There has been repeated discussion in the literature about whether researchers should or should not "correct for phylogeny" in a particular analysis by using approaches like PIC (Felsenstein 1985; Harvey et al. 1995; Rohlf 2006; Westoby et al. 2016) and a concern about "overcorrecting." The original formulation and explanatory figure for PIC has led to this terminology, which in turn may have led to some misunderstandings. Although it is true that without considering phylogeny, a comparative analysis of two species variables might be statistically invalid, it might be helpful to think of PIC (and PGLS) as incorporating the additional useful information that phylogeny provides into the estimation of the relationships between traits, rather than as a correction.

In the context of analyses like regression or ANOVA that can be used to understand trait correlations and adaptation, if there is a phylogenetic signal in the *residuals* from a model fit, then the resulting relationship derived is statistically invalid because the assumption of independent datapoints has been violated. Phylogenetic signal in any of the individual traits under investigation does not necessarily mean the residuals of the regression will have a phylogenetic structure, or vice versa. Revell (2010) provides a thorough explanation of this issue. If in doubt, rather than using the function `corBrownian` shown in Example Analysis 1, you can use `corPagel` to define the correlation structure. This will jointly estimate lambda, which is a measure of the phylogenetic signal in the residuals. As lambda gets closer to 0, the estimated coefficients in PGLS will converge on those estimated using OLS. Think about what that means in terms of model assumptions between OLS and PGLS. Just because "nonphylogenetic" methods like OLS do not incorporate information about evolutionary relationships does not mean they do not make assumptions about evolutionary change. In fact, the biological assumptions underlying OLS for comparative analyses are mathematically and conceptually equivalent to using a phylogeny, and assuming that it is a star phylogeny (i.e. all branches simultaneously diverge from a single node). Thus, both OLS and PGLS are based on models. It is important to keep in mind that all models are wrong, but some are more wrong than others. We advocate for using model goodness-of-fit tests like the Akaike information criterion (AIC) and Bayesian information criterion (BIC), or, where possible, going further and investigating model adequacy (see Section 7).

Example Analysis 1 – Regression

Required packages: `nlme; ape`

To load our example data, download it from github.com/daveyfwright/PCMsForPaleontologists and then follow the instructions in the file "Data_loading_script.R." To follow along, use the script in the file "full_script_PaleoPCM.R," or to see output without following along, see the file "example analyses.pdf." We include some of the output in the first two example analyses.

Is calyx shape correlated with fan density? To answer this, we could calculate the Pearson correlation coefficient, or perform an OLS regression.

Always plot the data, in this case `Shape` and `Density`, from the dataframe `allTraits`:

```
plot(allTraits$Shape, allTraits$Density,
     pch = 19, main = "",
     xlab = "Calyx shape ln(L/W)",
    ylab = "Fan density")
```

The standard test of correlation between two continuous variables is the Pearson correlation coefficient:

```
cor.test(allTraits$Shape, allTraits$Density)
```

Run this line of code and you should get the following output, which indicates shape and fan density are negatively correlated with p=0.02.

```
  data: allTraits$Shape and allTraits$Density
  t = -2.4878, df = 63, p-value = 0.01551
  alternative hypothesis: true correlation is not
  equal to 0
  95 percent confidence interval:
   -0.50606523 -0.05952433
  sample estimates:
        cor
   -0.2990814
```

The standard regression is OLS:

```
ols <- gls(Density ~ Shape,
           data = allTraits, method = "ML")
```